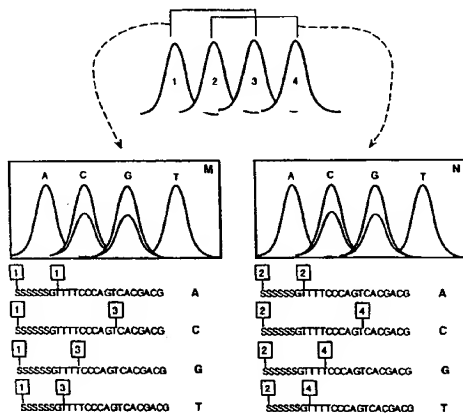




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12Q 1/68</b>		(11) International Publication Number: <b>WO 98/31834</b>
<b>A1</b>		(43) International Publication Date: 23 July 1998 (23.07.98)
(21) International Application Number: PCT/US97/22914		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 12 December 1997 (12.12.97)		
(30) Priority Data: 08/784,162 15 January 1997 (15.01.97) US		
(71) Applicant: INCYTE PHARMACEUTICALS, INC. (US/US); 3174 Porter Drive, Palo Alto, CA 94304 (US).		
(72) Inventor: JU, Jingyue; 562 Osprey Drive, Redwood City, CA 94065 (US).		
(74) Agent: FIELD, Bret, E.; Bozicevic & Reed LLP, Suite 200, 285 Hamilton Avenue, Palo Alto, CA 94301 (US).		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

(54) Title: SETS OF LABELED ENERGY TRANSFER FLUORESCENT PRIMERS AND THEIR USE IN MULTI COMPONENT ANALYSIS



## (57) Abstract

Sets of fluorescent energy transfer labels and methods for their use in multi component analysis, particularly nucleic acid enzymatic sequencing, are provided. In the subject sets, at least two of the labels are energy transfer labels comprising a common donor and acceptor fluorophore in energy transfer relationship separated by different distances and capable of providing distinguishable fluorescence emission patterns upon excitation at a common wavelength. The subject labels find particular use in a variety of multi component analysis applications, such as probes in FISH and multi array analyses, as well as primers in nucleic acid enzymatic sequencing applications.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## SETS OF LABELED ENERGY TRANSFER FLUORESCENT PRIMERS AND THEIR USE IN MULTI COMPONENT ANALYSIS

### INTRODUCTION

#### 5 Technical Field

The field of this invention is fluorescent labels, particularly fluorescently labeled primers for use in DNA sequencing applications.

#### Background of the Invention

10 Fluorescent labels find use in variety of different biological, chemical, medical and biotechnological applications. One example of where such labels find use is in polynucleotide sequencing, particularly in automated DNA sequencing, which is becoming of critical importance to large scale DNA sequencing projects, such as the Human Genome Project.

15 In methods of automated DNA sequencing, differently sized fluorescently labeled DNA fragments which terminate at each base in the sequence are enzymatically produced using the DNA to be sequenced as a template. Each group of fragments corresponding to termination at one of the four labeled bases are labeled with the same label. Thus, those fragments terminating in A are labeled with a first label, while those terminating in G, C and  
20 T are labeled with second, third and fourth labels respectively. The labeled fragments are then separated by size in an electrophoretic medium and an electropherogram is generated, from which the DNA sequence is determined.

As methods of automated DNA sequencing have become more advanced, of increasing interest is the use of sets of fluorescent labels in which all of the labels are excited  
25 at a common wavelength and yet emit one of four different detectable signals, one for each of the four different bases. Such labels provide for a number of advantages, including high fluorescence signals and the ability to electrophoretically separate all of the labeled fragments in a single lane of an electrophoretic medium which avoids problems associated with lane to lane mobility variation.

30 Although such sets of labels have been developed for use in automated DNA sequencing applications, heretofore the differently labeled members of such sets have each emitted at a different wavelength. Thus, conventional automated detection devices currently

employed in methods in which all of the enzymatically produced fragments or primer extension products are separated in the same lane must be able to detect emitted fluorescent light at four different wavelengths. This requirement can prove to be an undesirable limitation. More specifically, carrying out sequencing on vast numbers of different DNA templates simultaneously increases the number of different fragments and corresponding labels required. At the same time, there is a need for a reduction in the complexity of the detection device, *e.g.* a device which can operate with light detection at only two wavelengths is preferable.

It would therefore be desirable to develop sets of fluorescent labels capable of providing four distinguishable signals, where the number of wavelengths associated with the four different signals is less than the number of different labels, *e.g.* where four different labels provide signals comprising emitted light at from one to two wavelengths. With such sets one could either: (1) reduce the complexity of automated detector devices or (2) increase the throughput of detectors capable of detecting at four different wavelengths, thereby achieving sequencing two DNA templates, or the same double stranded templates from both the 5' and 3' end, simultaneously.

#### Relevant Literature

DNA sequencing is reviewed in Griffin & Griffin, *Appl. Biochem. Biotechnol.* (1993) 38:147-159. Fluorescence energy transfer labels and their use in DNA sequencing applications are described in Ju et al., *Nucleic Acids Res.* (1996) 24:1144-1148; Ju et al., *Nat. Med.* (1996) 2: 246-249; Ju et al., *Anal. Biochem.* (1995) 231: 131-140; Ju et al., *Proc. Natl. Acad. Sci. USA* (1995) 92: 4347-4351. Use of fluorescent energy transfer labels for non-DNA sequencing multi component analysis application is described in Wang et al., *Anal. Chem.* (1995) 67:1197-1203; Ziegler et al., *Genomics* (1992) 14: 1026-1031; and Repp et al., *Leukemia* (1995) 9: 210-215. Other references describing multi-component analysis applications include Schena et al., *Science* (1995) 270:467-469.

Other references of interest include U.S. Pat. Nos. 4,996,143 and 5,326,692, as well as Glazer and Streyer, *Biophys. J.* (1983) 43: 383-386, Huang et al., *Anal. Chem.* (1992) 64:2149-2154; Prober et al., *Science* (1987) 336-341; Smith et al., *Nature* (1986) 321: 674-679, Lu et al., *J. Chromat. A* (1994) 680: 497-501 and Ansorge et al., *Nucleic Acids Res.* (1987) 15: 4593-4603.

### SUMMARY OF THE INVENTION

Sets of fluorescent labels, particularly labeled primers, as well as methods for their use in multi component analysis, are provided. At least two of the labels of the subject sets comprise a common donor and acceptor fluorescer component in energy transfer relationship  
5 separated by different distances, such that the labels provide distinguishable fluorescent signals upon excitation at a common light wavelength. The subject sets of labels find use in a variety of applications requiring a plurality of distinguishable fluorescent labels, and find particular use as primers in nucleic acid enzymatic sequencing applications. Primers with the same labels which produce distinguishable emission patterns can be produced because  
10 energy transfer between the acceptor and donor fluorophores is a function of the separation distance between the acceptor and donor in the label. By changing the distance, different fluorescence emission patterns are obtained.

### BRIEF DESCRIPTION OF THE FIGURES

15 FIG. 1 shows the general labeling concept using four fluorescent molecules to generate at least eight fluorescent dye-labeled primers with distinguishable fluorescence emission patterns.

FIG. 2 shows the structure of the four primers labeled with two different fluorescent dyes, 6-carboxyfluorescein (FAM, F) as a donor and 6-carboxyrhodamine (ROX, R) as an acceptor.  
20 The numbers in the primer name indicate the intervening nucleotides between the donor and acceptor.

FIG. 3 shows that the fluorescence signal of the four fluorescent primers is sufficiently  
25 different to code for four nucleotides T (F6R), G (F13R), A (F16R), C (F16F). It also shows that the primer F16T using 6-carboxy tetramethyl rhodamine (TAMRA, T) to replace ROX (R) as an acceptor displays almost equal fluorescence signal intensity of F (blue) and T (black). The fluorescence signals shown are the electropherograms of the single base extension fragments from each primer obtained in the ABI four color fluorescent 377 DNA  
30 sequencer which has the appropriate filters to detect fluorescence signals from FAM (F  $\lambda_{em(max)}$  = 525 nm), ROX (R  $\lambda_{em(max)}$  = 605 nm) and TAMRA (T  $\lambda_{em(max)}$  = 580 nm).

FIG. 4 shows that the fluorescence intensity of the single base extension fragments from primer F6R (T fragments) and F13R (G fragments) due to energy transfer from F to R is much higher than that of the single base fragments generated with primer R15R (T fragments) which carries two ROX dyes but with same sequence as F6R and F13R. Same concentration of the primer and other sequencing reagents were used in the comparison.

FIG. 5 shows a small portion of the raw sequencing data in 2-color mode (FAM,  $F \lambda_{em(max)} = 525$  nm, blue; ROX,  $R \lambda_{em(max)} = 605$  nm, red) generated by primer F6R (T), F13R (G), F16R (A), F16F (C) and a cDNA clone which has a polyA tail at the 3' end. Sequences can be called by the color patterns of each peak.

FIG. 6 shows a large portion of the raw sequencing data (from nucleotide 30 to 130) in 2-color mode (FAM,  $F \lambda_{em(max)} = 525$  nm, blue; ROX,  $R \lambda_{em(max)} = 605$  nm, red) generated by primer F6R (T), F13R (G), F16R (A), F16F (C) and a cDNA clone which has a polyA tail at the 3' end. Sequences can be called by the color patterns of each peak. Samples were prepared using Thermo Sequenase Kit (Amersham LIFE SCIENCE) and run on a ABI 377 DNA sequencer with virtual filter A that detects the fluorescence signal from FAM and ROX.

FIG. 7 is a schematic of the Sanger enzymatic DNA sequencing method.

### DEFINITIONS

The term "fluorescent label" refers to a compound comprising at least one fluorophore bonded to a polymer.

The term "energy transfer fluorescent label" refers to a compound comprising at least two fluorophores in energy transfer relationship, where the fluorophores are bonded to a spacer component, e.g. a polymeric moiety, which separates the two fluorophores by a certain distance.

The term "enzymatic sequencing," "Sanger Method," "dideoxy technique," and "chain terminator technique," are used interchangeably herein to describe a method of sequencing DNA named after its main developer, F. Sanger. The technique uses a single-stranded DNA template, a short DNA primer and a polymerase enzyme to synthesize a

complementary DNA strand. The primer is first annealed to the single-stranded template and the reaction mixture is then split into four aliquots and deoxynucleoside triphosphates (dNTPs) plus a dideoxynucleoside triphosphate (ddNTP) are added such that each tube has a different ddNTP. The polymerase will incorporate a ddNTP opposite its complementary  
5 base on the template but no further dNTPs can be added as the ddNTP lacks a 3' hydroxyl group. The ratio of ddNTP to dNTP is such that the polymerase will terminate the growing DNA chain at all positions at which the ddNTP can be inserted and so a nested set of fragments (*i.e.* primer extension products) is formed which all have one end, the primer, in common. The fragments are labeled so that when the four reaction mixtures are  
10 electrophoresed through a polyacrylamide gel, a gel band pattern or ladder is formed from which the DNA sequence can be read directly. The process is shown schematically in Fig. 7.

The term "enzymatically produced" means produced at least in part as a result of an action of an enzyme, *e.g.* fragments of nucleotides are produced when an enzyme catalyzes a reaction whereby a larger sequences is cleaved into two or more fragments.

15 The term "primer" shall mean a polymer sequence which is complementary and capable of hybridizing to some part of a single stranded nucleotide sequence being sequenced which primer is used to initiate DNA synthesis *in vitro*.

#### DESCRIPTION OF THE SPECIFIC EMBODIMENTS

20 Sets of fluorescent labels, particularly sets of fluorescently labeled primers, and methods for their use in multi component analysis applications, particularly nucleic acid enzymatic sequencing applications, are provided. At least two of the label members of the set are energy transfer labels having a common donor and acceptor fluorophore separated by sufficiently different distances so that the two labels provide distinguishable fluorescent  
25 signals upon excitation at a common wavelength. In further describing the subject invention, the subject sets will first be described in greater detail followed by a discussion of methods for their use in multi component analysis applications.

Before the subject invention is further described, it is to be understood that the invention is not limited to the particular embodiments of the invention described below, as  
30 variations of the particular embodiments may be made and still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the

purpose of describing particular embodiments, and is not intended to be limiting. Instead, the scope of the present invention will be established by the appended claims.

It must be noted that as used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

The subject sets of fluorescent labels comprise a plurality of different types of labels, wherein each type of label in a given set is capable of producing a distinguishable fluorescent signal from that of the other types of labels in different sets. Labels in the different sets generate different signals, preferably, though not necessarily upon excitation at a common excitation wavelength. For DNA sequencing applications, the subject sets will comprise at least 2 different types of labels, and may comprise 8 or more different types of labels, where for many applications the number of different types of labels in the set will not exceed 6, and will usually not exceed four, where at least two of the different types of labels are energy transfer labels sharing a common donor and acceptor fluorescer, as described in greater detail below. For other applications, such as fluorescence in situ hybridization (FISH), substantially more than 8 labels are ideal so that multiple targets can be analyzed.

The distinguishable signals generated by the "at least two energy transfer labels" will at least comprise the intensity of emitted light at one to two wavelengths. Preferably, the distinguishable signals produced by the "at least two energy transfer labels" will comprise distinguishable fluorescence emission patterns, which patterns are generated by plotting the intensity of emitted light from differently sized fragments at two wavelengths with respect to time as differently labeled fragments move relative to a detector, which patterns are known in the art as electropherograms. For analyses not based on electrophoresis, such as micro-array chip based assays, different targets tagged with a specific label can be differentiated from each other by the unique fluorescence patterns. For example, in one type of label of a set the intensity of emitted light at a first wavelength may be twice that of the intensity of emitted light at a second wavelength and in the second label the magnitude of the intensities of light emitted at the two wavelengths may be reversed, or light may be emitted at only one intensity. The different patterns are generated by varying the distance between the donor and acceptor. These patterns emitted from each of these labels are thus distinguishable.



The subject sets will comprise a plurality of different types of fluorescent labels, where at least two of the labels and usually all of the labels are energy transfer labels which comprise at least one acceptor fluorophore and at least one donor fluorophore in energy transfer relationship, where such labels may have more complex configurations, such as multiple donors and/or multiple acceptors, e.g. donor 1, acceptor 1 and acceptor 2. Critical to the subject sets is that at least two of the labels of the sets have common donor and acceptor fluorophores, where the only difference between the labels is the distance between these common acceptor and donor fluorophores. Thus, for sets of labels in which each label comprises a single donor and a single acceptor, at least one of the energy transfer labels will have a donor fluorophore and acceptor fluorophore in energy transfer relationship separated by a distance  $x$  and at least one of the energy transfer labels will comprise the same donor and acceptor fluorophores in energy transfer relationship separated by a different distance  $y$ , where the distances  $x$  and  $y$  are sufficiently different to provide for distinguishable fluorescence emission patterns upon excitation at a common wavelength, as described above. In those sets comprising a third label having the same donor and acceptor fluorophores as the first and second label, the distance  $z$  between the donor and acceptor fluorophore will be sufficiently different from  $x$  and  $y$  to ensure that the third label is capable of providing a distinguishable fluorescence emission pattern from the first and second labels. Thus, in a particular set of labels, one may have a plurality of labels having the same donor and acceptor fluorophores, where the only difference among the labels is the distance between the donor and acceptor fluorophores. To ensure that different types of labels of a set having common donor and acceptor fluorophores yield distinguishable fluorescence emission patterns, the distances between the donor and acceptor fluorophores will differ by at least about 5 %, usually by at least about 10 % and more usually by at least about 20 % and will generally range from about 4 to 200 Å, usually from about 12 to 100 Å and more usually from about 15 to 80 Å, where the minimums in such distances are determined based on currently available detection devices and may be reduced as detection technology becomes more sensitive, therefore more distinct labels can be generated.

In one preferred embodiment, at least a portion of, up to and including all of, the labels of the subject sets will comprise a donor and acceptor fluorophore component in energy transfer relationship and covalently bonded to a spacer component, *i.e.* energy transfer labels. Thus, one could have a set of a plurality of labels in which only two of the labels comprise

the above mentioned donor and acceptor fluorescer components and the remainder of the labels comprise a single fluorescer component. Preferably, however, all of the labels will comprise a donor and acceptor fluorescer component. Generally, for one donor and one acceptor ET systems, if a set comprises  $n$  types of energy transfer labels, the number of different types of acceptor fluorophores present in the energy transfer labels of the set will not exceed  $n-1$ . Thus, if the number of different types of energy transfer labels in the set is four, the number of different acceptor fluorophores in the set will not exceed 3, and will usually not exceed 2.

In other preferred embodiments, additional combinations of labels are possible.

Thus, in a set of labels, two of the labels could be energy transfer labels sharing common donor and acceptor fluorophores separated by different distances and the remaining labels could be additional energy transfer labels with different donor and/or acceptor fluorophores, non-energy transfer fluorescent labels, and the like.

In the energy transfer labels of the subject sets, the spacer component to which the fluorescer components are covalently bound will typically be a polymeric chain or other chemical moiety capable of acting as a spacer for the donor and acceptor fluorophore components, such as a rigid chemical moiety, such as chemicals with cyclic ring or chain structures which can separate the donor and acceptor and which also can be incorporated with an active group for attaching to the targets to be analyzed, where the spacer component will generally be a polymeric chain, where the fluorescer components are covalently bonded through linking groups to monomeric units of the chain, where these monomeric units of the chain are separated by a plurality of monomeric units sufficient so that energy transfer can occur from the donor to acceptor fluorescer components. The polymeric chains will generally be either polynucleotides, analogues or mimetics thereof; or peptides, peptide analogues or mimetics thereof, *e.g.* peptoids. For polynucleotides, polynucleotide analogues or mimetics thereof, the polymeric chain will generally comprise sugar moieties which may or may not be covalently bonded to a heterocyclic nitrogenous base, *e.g.* adenine, guanine, cytosine, thymine, uracil *etc.*, and are linked by a linking group. The sugar moieties will generally be five membered rings, *e.g.* ribose, or six membered rings, *e.g.* hexose, with five membered rings such as ribose being preferred. A number of different sugar linking groups may be employed, where illustrative linking groups include phosphodiester, phosphorothioate, methylene(methyl imino)(MMI), methophosphonate, phosphoramadite,

guanidine, and the like. *See* Matteucci & Wagner, *Nature* (1996) Supp 84: 20-22. Peptide, peptide analogues and mimetics thereof suitable for use as the polymeric spacer include peptoids as described in WO 91/19735, the disclosure of which is herein incorporated by reference, where the individual monomeric units which are joined through amide bonds may  
5 or may not be bonded to a heterocyclic nitrogenous base, *e.g.* peptide nucleic acids. *See* Matteucci & Wagner *supra*. Generally, the polymeric spacer components of the subject labels will be peptide nucleic acid, polysugarphosphate as found in energy transfer cassettes as described in PCT/US96/13134, the disclosure of which is herein incorporated by reference, and polynucleotides as described in PCT/US95/01205, the disclosure of which is  
10 herein incorporated by reference.

Both the donor and acceptor fluorescer components of the subject labels will be covalently bonded to the spacer component, *e.g.* the polymeric spacer chain, through a linking group. The linking group can be varied widely and is not critical to this invention. The linking groups may be aliphatic, alicyclic, aromatic or heterocyclic, or combinations  
15 thereof. Functionalities or heteroatoms which may be present in the linking group include oxygen, nitrogen, sulfur, or the like, where the heteroatom functionality which may be present is oxy, oxo, thio, thiono, amino, amido and the like. Any of a variety of the linking groups may be employed which do not interfere with the energy transfer and gel electrophoresis, which may include purines or pyrimidines, particularly uridine, thymidine,  
20 cytosine, where substitution will be at an annular member, particularly carbon, or a side chain, *e.g.* methyl in thymidine. The donor and/or fluorescer component may be bonded directly to a base or through a linking group of from 1 to 6, more usually from 1 to 3 atoms, particularly carbon atoms. The linking group may be saturated or unsaturated, usually having not more than about one site of aliphatic unsaturation.

Though not absolutely necessarily, generally for DNA sequencing applications at least one of the donor and acceptor fluorescer components will be linked to a terminus of the polymeric spacer chain, where usually the donor fluorescer component will be bonded to the terminus of the chain, and the acceptor fluorescer component bonded to a monomeric unit  
25 internal to the chain. For labels comprising polynucleotides, analogues or mimetics thereof as the polymeric chain, the donor fluorescer component will generally be at the 5' terminus  
30 of the polymeric chain and the acceptor fluorescer component will be bonded to the

polymeric chain at a position 3' position to the 5' terminus of the chain. For other applications, such as FISH, a variety of labeling approaches are possible.

The donor fluorescer components will generally be compounds which absorb in the range of about 300 to 900 nm, usually in the range of about 350 to 800 nm, and are capable of transferring energy to the acceptor fluorescer component. The donor component will have a strong molar absorbance co-efficient at the desired excitation wavelength, desirably greater than about  $10^4$ , preferably greater than about  $10^5 \text{ cm}^{-1}\text{M}^{-1}$ . The molecular weight of the donor component will usually be less than about 2.0 kD, more usually less than about 1.5 kD. A variety of compounds may be employed as donor fluorescer components, including fluorescein, phycoerythrin, BODIPY, DAPI, Indo-1, coumarin, dansyl, cyanine dyes, and the like. Specific donor compounds of interest include fluorescein, rhodamine, cyanine dyes and the like.

Although the donor and acceptor fluorescer component may be the same, e.g. both may be FAM, where they are different the acceptor fluorescer moiety will generally absorb light at a wavelength which is usually at least 10 nm higher, more usually at least 20 nm or higher, than the maximum absorbance wavelength of the donor, and will have a fluorescence emission maximum at a wavelength ranging from about 400 to 900 nm. As with the donor component, the acceptor fluorescer component will have a molecular weight of less than about 2.0 kD, usually less than about 1.5 kD. Acceptor fluorescer moieties may be rhodamines, fluorescein derivatives, BODIPY and cyanine dyes and the like. Specific acceptor fluorescer moieties include FAM, JOE, TAM, ROX, BODIPY and cyanine dyes.

The distance between the donor and acceptor fluorescer components will be chosen to provide for energy transfer from the donor to acceptor fluorescer, where the efficiency of energy transfer will be from 20 to 100 %. Depending on the donor and acceptor fluorescer components, the distance between the two will generally range from 4 to 200 Å, usually from 12 to 100 Å and more usually from 15 to 80 Å, as described above.

For the most part the labels of the subject sets will be described by the following formula:

30



wherein:

D is the donor fluorescer component, which may consist of more than two different donors separated by a spacer;

N is the spacer component, which may be a polymeric chain or rigid chemical moiety, where when N is a polymeric spacer that comprises nucleotides, analogues or mimetics thereof, the number of monomeric units in N will generally range from about 1 to 50, usually from about 4 to 20 and more usually from about 4 to 16;

A is the acceptor fluorescer component, which may consist of more than two different acceptors separated by a spacer; and

X is optional and is generally present when the labels are incorporated into oligonucleotide primers, where X is a functionality, *e.g.* an activated phosphate group, for linking to a mono- or polynucleotide, analogue or mimetic thereof, particularly a deoxyribonucleotide, generally of from 1 to 50, more usually from 1 to 25 nucleotides.

For sets to be employed in nucleic acid enzymatic sequencing in which the labels are to be employed as primers, the labels of the subject sets will comprise either the donor and acceptor fluorescer components attached directly to a hybridizing polymeric backbone, *e.g.* a polynucleotide, peptide nucleic acid and the like, or the donor and acceptor fluorescer components will be present in an energy transfer cassette attached to a hybridizable component, where the energy transfer cassette comprises the fluorescer components attached to a non-hybridizing polymeric backbone, *e.g.* a universal spacer. *See* PCT/US96/13134 and Ju *et al.*, Nat. Med. (1996) *supra*, the disclosures of which are herein incorporated by reference. The hybridizable component will typically comprise from about 8 to 40, more usually from about 8 to 25 nucleotides, where the hybridizable component will generally be complementary to various commercially available vector sequences such that during use, synthesis proceeds from the vector into the cloned sequence. The vectors may include single-stranded filamentous bacteriophage vectors, the bacteriophage lambda vector, pUC vectors, pGEM vectors, or the like. Conveniently, the primer may be derived from a universal primer, such as pUC/M13,  $\lambda$ gt10,  $\lambda$ gt11, and the like, (*See* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd ed., CSHL, 1989, Section 13), where the universal primer will have been modified as described above, *e.g.* by either directly attaching the donor and acceptor fluorescer components to bases of the primer or by attaching an energy transfer cassette comprising the fluorescer components to the primer.

Sets of preferred energy transfer labels comprising donor and acceptor fluorescers covalently attached to a polynucleotide backbone in the above D-N-A format include: (1) F6R, F13R, F16R and F16F; where different formats can employed as long as the four primers display distinct fluorescence emission patterns.

- 5           The fluorescent labels of the subject sets can be readily synthesized according to known methods, where the subject labels will generally be synthesized by oligomerizing monomeric units of the polymeric chain of the label, where certain of the monomeric units will be covalently attached to a fluorescer component.

- 10           The subject sets of fluorescent labels find use in applications where at least two components of a sample or mixture of components are to be distinguishably detected. In such applications, the set will be combined with the sample comprising the to be detected components under conditions in which at least two of the components of the sample if present at all will be labeled with first and second labels of the set, where the first and second labels of the set comprise the same donor and acceptor fluorescer components which are  
15           separated by different distances. Thus, a first component of the sample is labeled with a first label of the set comprising donor and acceptor fluorescer components separated by a first distance X. A second component of the sample is labeled with a second label comprising the same donor and fluorescer components separated by a second distance Y, where X and Y are as described above. The labeled first and second components, which may or may not have  
20           been separated from the remaining components of the sample, are then irradiated by light at a wavelength capable of a being absorbed by the donor fluorescer components, generally at a wavelength which is maximally absorbed by the donor fluorescer components. Irradiation of the labeled components results in the generation of distinguishable fluorescence emission patterns from the labeled components, a first fluorescence emission pattern generated by the  
25           first label and second pattern being attributable to the second label. The distinguishable fluorescence emission patterns are then detected. Applications in which the subject labels find use include a variety of multicomponent analysis applications in which fluorescent labels are employed, including FISH, micro-array chip based assays where the labels may be used as probes which specifically bind to target components, DNA sequencing where the  
30           labels may be present as primers, and the like.

The subject sets of labels find particular use in polynucleotide enzymatic sequencing applications, where four different sets of differently sized polynucleotide fragments

terminating at a different base are generated (with the members of each set terminating at the same base) and one wishes to distinguish the sets of fragments from each other. In such applications, the sets will generally comprise four different labels which are capable of acting as primers for enzymatic extension, where at least two of the labels will be energy transfer labels comprising differently spaced common donor and acceptor fluorescer components that are capable of generating distinguishable fluorescence emission patterns upon excitation at a common wavelength of light. Using methods known in the art, a first set of primer extension products all ending in A will be generated by using a first of the labels of the set as a primer. Second, third and fourth sets of primer extension products terminating in G, C and T will be also be enzymatically produced. The four different sets of primer extension products will then be combined and size separated, usually in an electrophoretic medium. The separated fragments will then be moved relative to a detector (where usually either the fragments or the detector will be stationary). The intensity of emitted light from each labeled fragment as it passes relative to the detector will be plotted as a function of time, *i.e.* an electropherogram will be produced. Since, the labels of the subject sets will generally emit light in only two wavelengths, the plotted electropherogram will comprise light emitted in two wavelengths. Each peak in the electropherogram will correspond to a particular type of primer extension product (*i.e.* A, G, C or T), where each peak will comprise one of four different fluorescence emission patterns. To determine the DNA sequence, the electropherogram will be read, with each different fluorescence emission pattern related to one of the four different bases in the DNA chain.

Where desired, two sets of labels according to the subject invention may be employed, where the distinguishable fluorescence emission patterns produced by the labels in the first set will comprise emissions at a first and second wavelength and the patterns produced by the second set of labels will comprise emissions at a third and fourth wavelength. By using two such sets in conjunction with one another, one could detect primer extension products produced from two different template DNA strands at essentially the same time in a conventional four color detector, thereby doubling the throughput of the detector.

30

The subject sets of labels may be sold in kits, where the kits may or may not comprise additional reagents or components necessary for the particular application in which

the label set is to be employed. Thus, for sequencing applications, the subject sets may be sold in a kit which further comprises one or more of the additional requisite sequencing reagents, such as polymerase, nucleotides, dideoxynucleotides and the like.

5           The following examples are offered by way of illustration and not by way of limitation. The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject sets of fluorescent labels.

10 EXPERIMENTAL

### A. Design and synthesis of the fluorescent primers.

An example of a general labeling scheme using the energy transfer concept to generate at least eight fluorescent primers from four fluorescent dyes is described in Figure

1. To demonstrate the practicality of the labeling approach, two fluorescent dyes 6-carboxyfluorescein (FAM,  $\lambda_{\text{em(max)}} = 525 \text{ nm}$ ) as a donor and 6-carboxy-X-rhodamine (ROX,  $\lambda_{\text{em(max)}} = 605 \text{ nm}$ , red) as an acceptor are chosen to generate four fluorescent oligonucleotide primers, which are subsequently used for DNA sequencing on a cDNA clone. The structures of the fluorescent primers are presented in Figure 2.

Oligodeoxynucleotides (25-bases long) with the sequence

20 5'-TTTTTTTTTTTTTTTTTTTTTAC-3' (SEQ ID NO:01) were synthesized with donor-acceptor fluorophore pairs separated by different distances. The 25-mer contains a modified base introduced by the use of 5'-dimethoxytrityl-5-[N-(trifluoroacetylaminohexyl)-3-acrylimido]-2'-deoxyuridine, 3'-[(2-cyano-ethyl)-(N,N-diisopropyl)]-phosphoramidite (Amino-Modifier C6 dT, Glen Research, Sterling, VA) which has a protected primary amine linker arm. The donor dye was attached to the 5' end of the oligomer, and the acceptor dye was attached to the primary amine group on the modified base. The primers are synthesized and purified according to the published procedure (Ju, J., Ruan, C., Fuller, C. W. Glazer, A. N. and Mathies, R. A. (1995) Proc. Natl. Acad. Sci. USA 92, 4347-4351). The ET primers are named using the abbreviation D-N-A, where D is the donor, A is the acceptor, and N is the number of intervening nucleotides between D and A. In all the primers prepared, 6-carboxyfluorescein (FAM, F, with fluorescence emission maximum at 525 nm) is selected as a common donor, and 6-carboxy-X-rhodamine (ROX, R, with fluorescence emission



maximum at 605 nm) is selected as an acceptor, except in one example where N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA, T, with fluorescence emission maximum at 580 nm) is chosen as an acceptor as shown in Figure 3.

Five fluorescent primers with their unique fluorescence signal patterns are shown in Figure 3. For primer F6R, the energy transfer efficiency from donor F to R is higher than 90%, therefore it only displays a dominant red color from acceptor R. For primer F13R, the ET efficiency is less than that in F6R, therefore F13R not only displays a high red signal from the acceptor R and also a blue signal from F with intensity approximately 40% of the red signal. For primer F16R, the ET efficiency is even less than that in F13R, which results in an approximately equal signal intensity from R (red) and F (blue). For primer F16F carrying two FAM molecules, the fluorescence signal is dominated by blue. For primer F16T that uses TAMRA (T) as an acceptor, the fluorescence signal from F (blue) is almost in equal intensity as T (Black). It is clear from these examples that at least five different fluorescent labels are generated using only three dyes. With two dyes FAM and ROX, four fluorescent primers are generated that have sufficiently different fluorescence signals to code for the for DNA sequencing fragments ended with nucleotides T (F6R), G (F13R), A (F16R) and C (F16F). These four primers are then chosen for evaluation in DNA sequencing using a cDNA clone with a polyA tail which can be primed by the designed primers.

## 20 B. DNA Sequencing procedure.

Sequencing was performed using a cDNA clone labeled Incyte clone 1 shown below and Thermo Sequenase sequencing kit (Amersham Life Science) on an ABI 377 sequencer.

### INCYTE CLONE 1

(The italic sequence is the one shown in Fig. 6.)

25 CNCGNCCAGT GAATTGTAAT ACGACTCACT ATAGGGCGAA TTGGGTACCG  
GGCCCCCCT CGAGTTTTTT TTTTTTTTTT TTTACATGAA GGCAATTTAT  
TAACAGAAAA TATTTTGAGG AATCTTGTTT ACAGACGGCG ACCACGGCGA  
CCCCCCTTCC TGCAGTGCT GTCAGAGGGG ATGGGGGTGA CATCCTCAAT  
CCGCCCAGTC TTCATACCCG AGCGGGCAAG GGCTCTGAGG GCCGACTGGG  
30 CCCAGGTCC AGGGGTCTTG GTCCTATTTT CTCCTGTGGC CCGAGTTTG  
(SEQ ID NO:02)

Four reactions were run, one for each dye/ddNTP combination with 0.2 pmole of the appropriate primer. The reactions containing ddCTP were run with the F16F primer, ddATP with the F16R primer, ddGTP with the F13R, and ddTTP with the F6R primer. Fifteen cycles of 94°C for 20 seconds, 47°C for 40 seconds and 68°C for 60 seconds were carried out for the sequencing reaction mixture and then cooled to 4°C. The four reaction mixtures for each sequence were then combined into one vial and 50 µl of 100% ethanol were added to precipitate the DNA fragments. The DNA was precipitated by centrifugation for 30 min at 4°C and then washed once with 70% ethanol. The precipitated DNA was vacuum dried, and resuspended in 4 µl of deionized formamide containing 8.3 mM EDTA and heated at 95°C for 2 min. The denatured DNA was loaded on a 4% polyacrylamide 7 M urea denaturing gel mounted in the instrument. Electrophoresis was conducted for 3.5 hours using 1X Tris-borate-EDTA buffer.

C. DNA sequencing results with the four fluorescent primers.

Figure 4 shows that the fluorescence intensity of the single base extension fragments from primer F6R (T fragments) and F13R (G fragments) due to energy transfer from F to R is much higher than that of the single base fragments generated with primer R15R (T fragments) which carries two ROX dyes but with same sequence as F6R and F13R. The same concentration of the primer and other sequencing reagents were used in the comparison. A small portion of the DNA sequencing raw data in a two color mode sampled from FAM and ROX using primer F6R, F13R, F16R and F16F on an ABI 377 DNA sequencer is shown in Figure 5. From this raw data, sequences can be determined by the color ratio of the peak in the electropherograms. Figure 6 shows a large portion of the raw sequencing data (from nucleotide 30 to 130) in 2-color mode generated by primer F6R (T), F13R (G), F16R (A), F16F (C) and a cDNA clone which has a polyA tail at the 3' end. Sequences can be called by the color patterns of each peak without applying any mobility shift correction on the raw data. For example, when the blue and red signals under one peak have almost the same intensity, the peak is assigned as an A; when only a dominant blue signal is seen in a peak, it was assigned as a C; when red signal is slightly higher than the blue signal in a peak, it was assigned as a G; when the red signal is much higher than the blue signal in a peak, it was assigned as a T.

It is clear from the experimental data that with two fluorescent dyes, using energy transfer concepts which offer higher fluorescence signals, four fluorescent primers can be generated with sufficiently different fluorescence signal patterns for sequencing DNA successfully. With two additional different fluorescent molecules, using the same principle presented, another four fluorescent primers can be constructed. Thus, with a DNA sequencer equipped with the appropriate 4-color filters, two sets of DNA sequencing samples can be analyzed simultaneously, doubling the sequencing throughput. These sets of unique fluorescent labels constructed with the concepts presented will find wide applications in other multiple component analysis projects.

It is evident from the above results and discussion that the sets of a labels of the subject invention provide for a number of advantages. For example, where the sets of labels are employed in DNA sequencing, one can employ a detector capable of detecting at only two very well separated wavelengths. Therefore, the detector can sample a large portion of the fluorescence signals, providing higher sensitivity for detection and increasing readlength of the sequence. Alternatively, with conventional detectors comprising four different wavelength detectors, one can effectively double the throughput obtainable with these detectors by sequencing two different strands with two sets of labels according to the subject invention. Thus, the subject sets of labels represent a significant contribution to the art.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

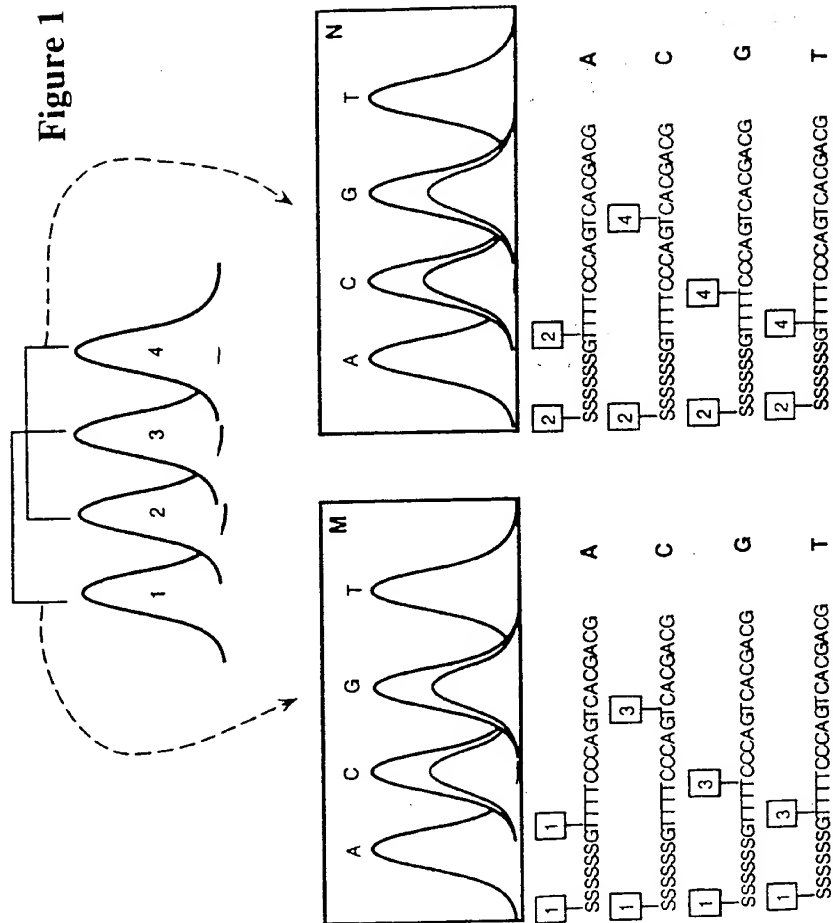
Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. A set of at least 2 different fluorescent labels, said set comprising:  
a first energy transfer fluorescent label comprising at least one donor fluorophore and  
at least one acceptor fluorophore in energy transfer relationship separated by a distance x;  
5 and  
a second energy transfer fluorescent label comprising said at least one donor  
fluorophore and said at least one acceptor fluorophore in energy transfer relationship  
separated by a distance y;  
wherein x and y are sufficiently different so that said first and second fluorescent  
10 labels provide distinguishable fluorescent signals.
2. The set according to Claim 1, wherein said acceptor fluorophore of said first and  
second labels is selected from the group consisting of TAM, JOE, FAM, ROX, BODIPY and  
cyanine dyes.  
15
3. The set according to Claim 1 or 2, wherein said set comprises n distinct energy  
transfer labels, wherein the total number of different acceptor fluorophores among said n  
distinct energy transfer labels does not exceed n-1.
- 20 4. The set according to Claims 1, 2 or 3 wherein said energy transfer labels comprise an  
energy transfer cassette.
5. The set according to any of the preceding claims, wherein all of the labels of said set  
are excited at a common wavelength.  
25
6. The set according to any of the preceding claims, wherein said labels are primers.
7. The set according to any of the preceding claims, wherein said donor fluorophore is  
FAM.  
30
8. The set according to any of the preceding claims, wherein said acceptor fluorophore  
is ROX.

9. A method of sequencing a DNA molecule, said method comprising:  
enzymatically producing four sets of primer extension products using said DNA molecule as a template, wherein each of said sets comprises a family of different sized primer extension products terminating at the same base, and further wherein each of said sets  
5 comprises the same fluorescent label, wherein the fluorescent labels are labels according to any of the preceding claims,  
size separating said primer extension products;  
detecting said size separated primer extension products; and  
determining said sequence from said detected size separated primer extension  
10 products.
10. A kit for use in DNA sequencing, said kit comprising:  
a set of four fluorescently labeled oligonucleotide primers, wherein at least two of said oligonucleotide primers have a common donor and acceptor fluorophore in energy  
15 transfer relationship and are separated by different distances.

1/7



2/7

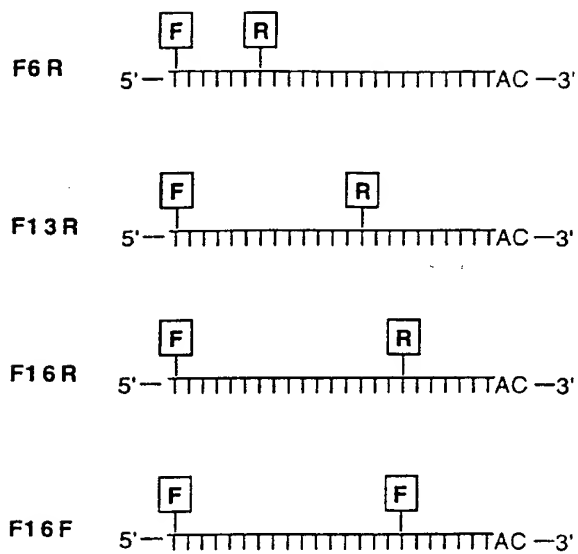


Figure 2

3/7

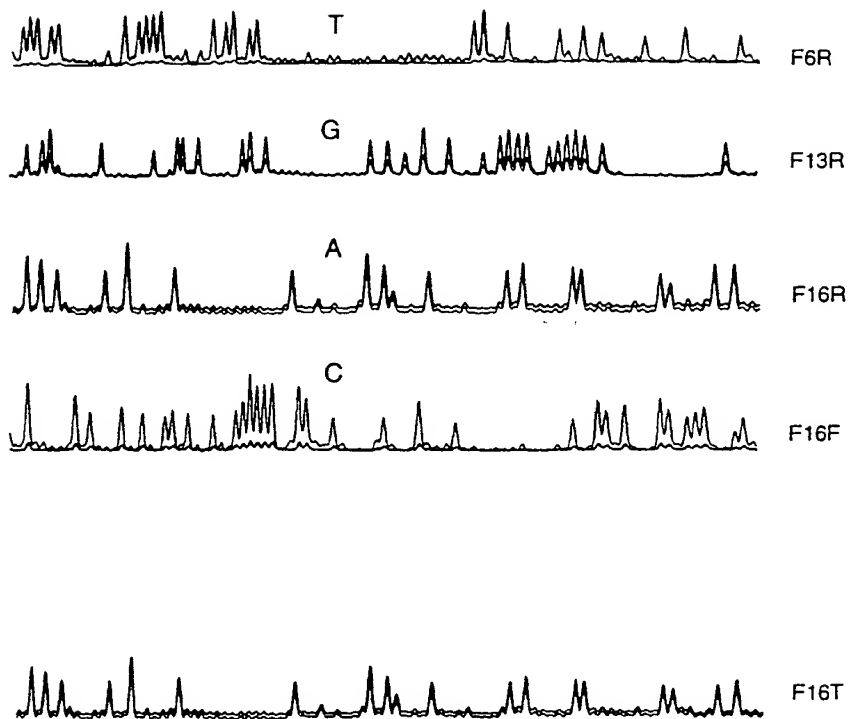


Figure 3



4/7

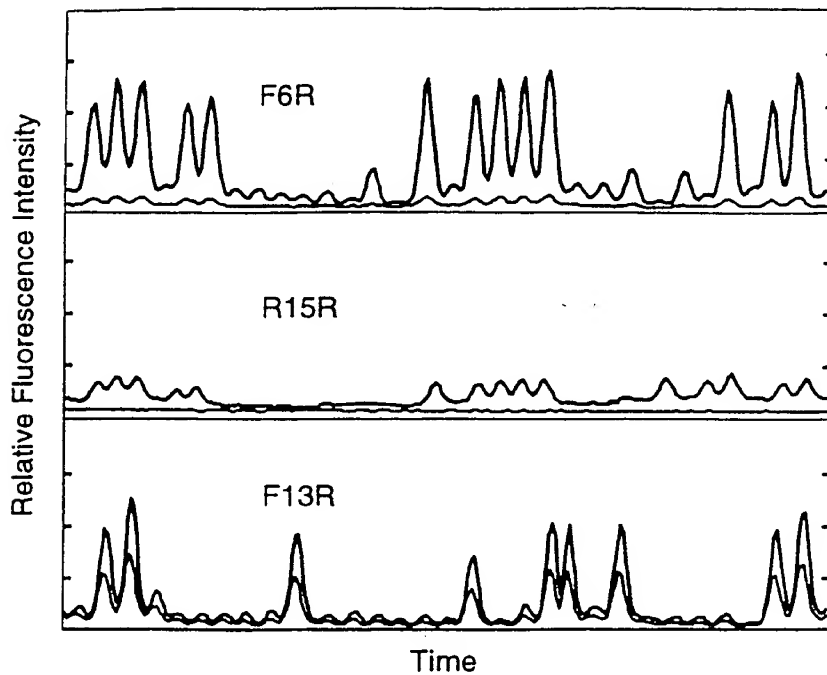


Figure 4

5/7

ACGGCGACCCCCCTTCCTGCG  
50 60 70

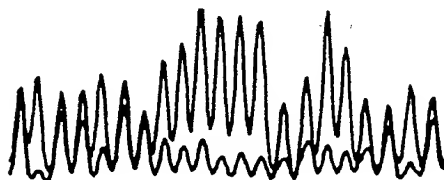


Figure 5

6/7

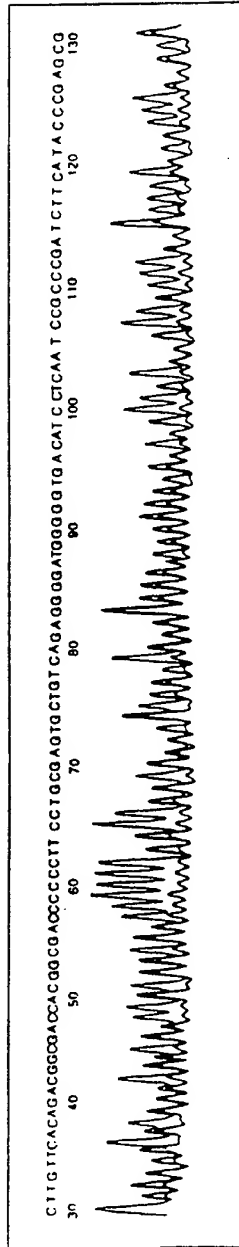


Figure 6



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 97/22914

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JU ET AL.: "ENERGY TRANSFER PRIMERS: A NEW FLUORESCENCE LABELING PARADIGM FOR DNA SEQUENCING AND ANALYSIS" NATURE MEDICINE, vol. 2, no. 2, February 1996, pages 246-249, XP002063932 cited in the application see the whole document ---	1-10
X	WO 95 21266 A (UNIV CALIFORNIA) 10 August 1995 see abstract see page 2, line 23 - page 3, line 3 see page 6, line 12 - line 21 --- -/--	1,2,5-10

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

30 April 1998

Date of mailing of the international search report

17. 06. 1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Ceder, O

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/22914

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>JU J ET AL: "FLUORESCENCE ENERGY TRANSFER DYE-LABELED PRIMERS FOR DNA SEQUENCING AND ANALYSIS"            PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,            vol. 92, May 1995, WASHINGTON, DC,            pages 4347-4351, XP002052420            cited in the application            see the whole document            ---</p>	1,2,5-10
X	<p>JU J ET AL: "DESIGN AND SYNTHESIS OF FLUORESCENCE ENERGY TRANSFER DYE-LABELED PRIMERS AND THEIR APPLICATION FOR DNA SEQUENCING AND ANALYSIS"            ANALYTICAL BIOCHEMISTRY,            vol. 231, no. 1, 10 October 1995, ORLANDO, FL,            pages 131-140, XP000536518            cited in the application            see the whole document            ---</p>	1,2,5-10
A	<p>JU ET AL.: "CASSETTE LABELING FOR FACILE CONSTRUCTION OF ENERGY TRANSFER FLUORESCENT PRIMERS"            NUCLEIC ACIDS RESEARCH,            vol. 24, no. 6, 1996,            pages 1144-1148, XP002063933            cited in the application            see the whole document            -----</p>	1-10

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. l. Application No

PCT/US 97/22914

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9521266 A	10-08-95	US 5654419 A	05-08-97
		AU 1736795 A	21-08-95
		CA 2182516 A	10-08-95
		DE 19581489 T	02-01-97
		DE 29521620 U	13-11-97
		EP 0743987 A	27-11-96
		JP 9508525 T	02-09-97
		US 5707804 A	13-01-98
		US 5688648 A	18-11-97
-----			